



Discriminant Analysis

All the suggested texts cover DA; also, [here's a nice web overview](#) and a quick google search will turn up dozens of sites.

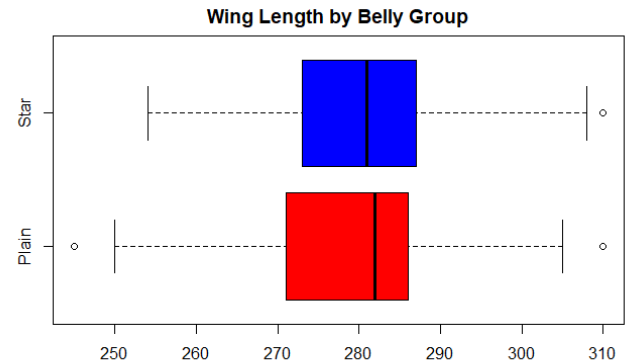
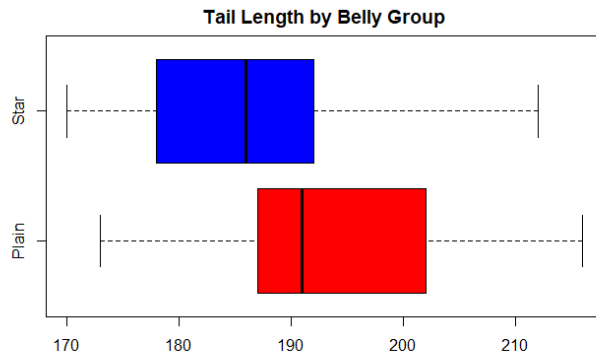


Example: Sneetches. The Sneetches were originally divided among those with stars upon thars and those with no stars upon thars (and there was great discrimination against those without stars). A Sylvester McMonkey McBean comes along and makes a mess of things with his star-on and star-off machines. Soon, nobody can tell which Sneetches were the original star-bellied sneetches. However, previous Sneetch medical records have tail and wing lengths in (mm) for both kinds of sneetches. A sample of 45 each of star bellied and plain bellied is as follows:

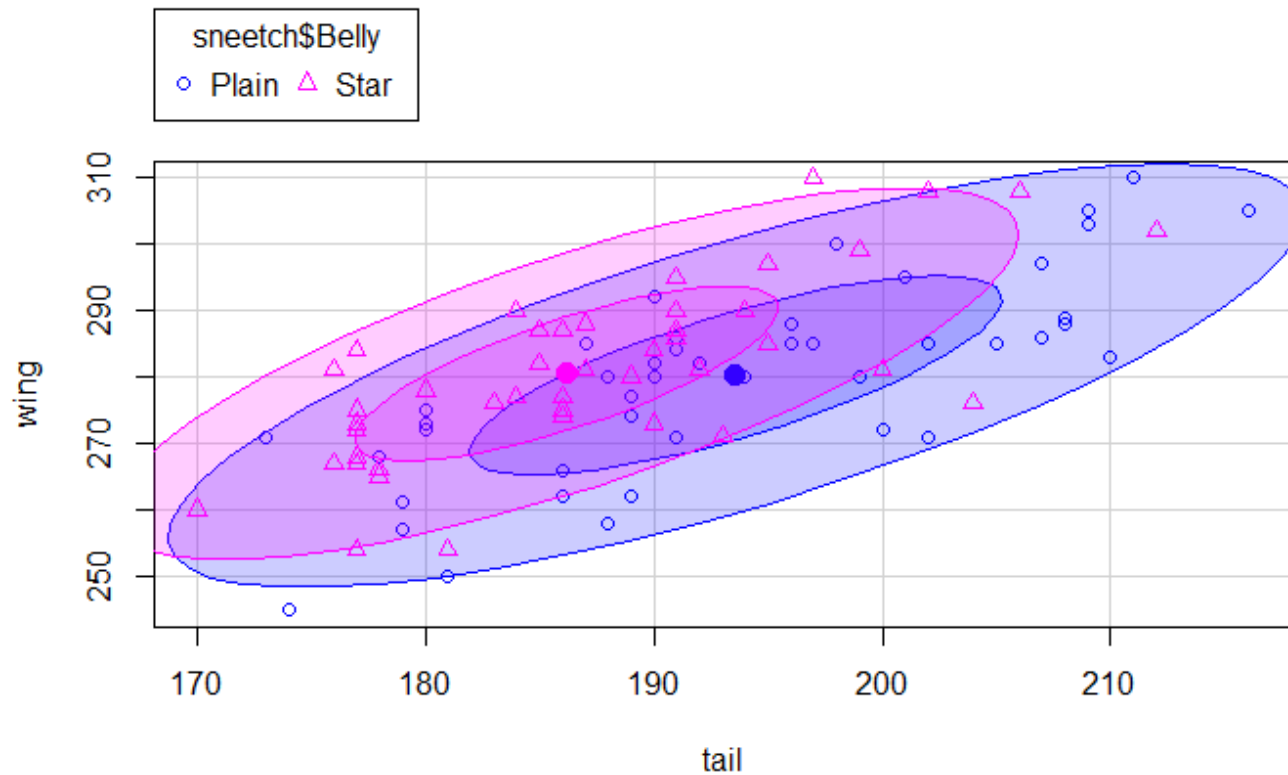


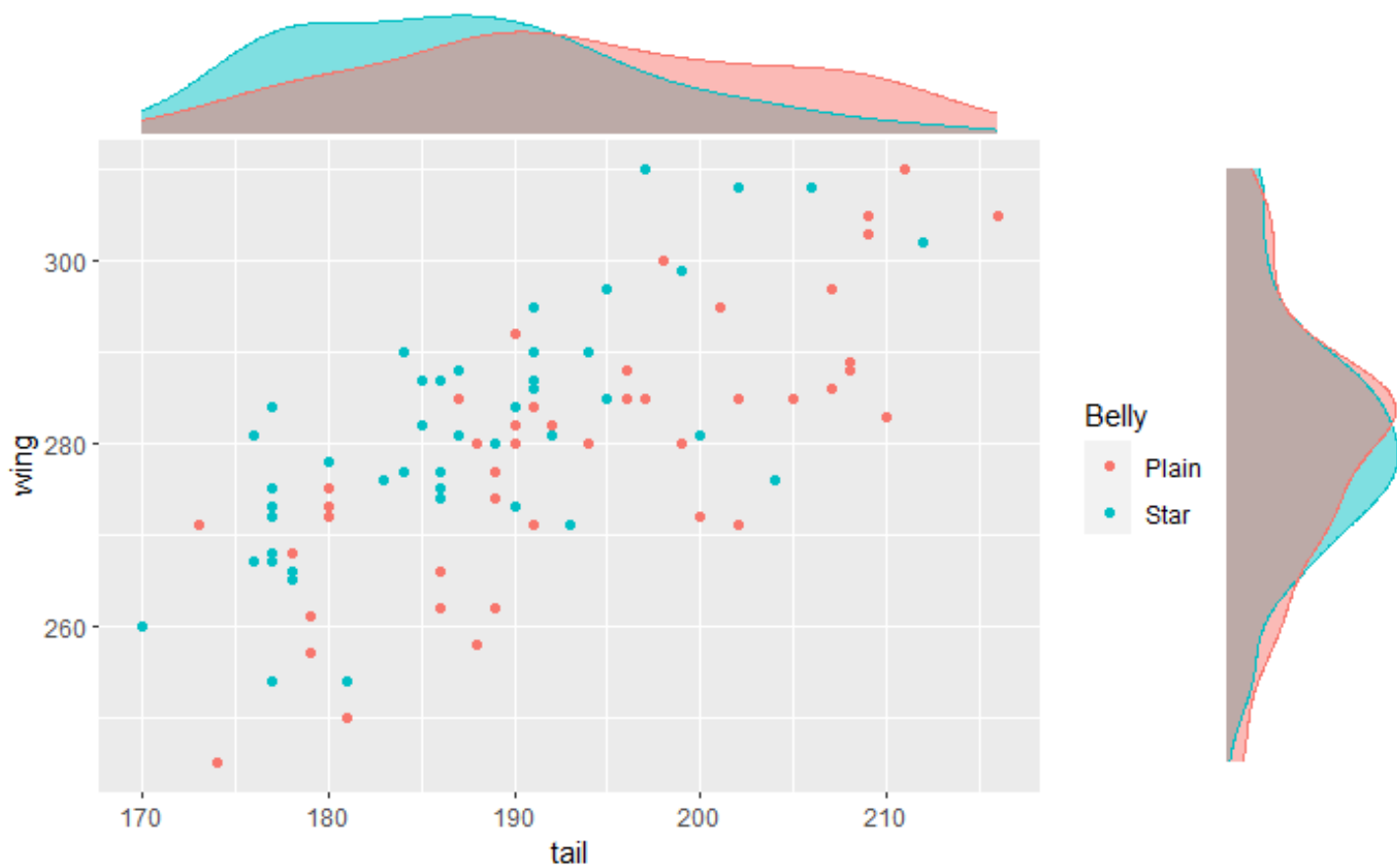


Boxplots of each variable indicate there is more difference between bellies in Tail length than in Wing length:



Here are a few more good ways of visualizing what's going on (code for these plots in R is on the CANVAS page):





DA Goal 1: Discrimination - identify variables that 'best' discriminate between two known groups.



We could do a t-test between belly groups on either wing length or tail length. However, is there a way to use the information of both variables simultaneously?

DA Goal 2: Find a function of identified variables to create a new variable which parsimoniously represents differences between groups.



Thanks to McBean, it's difficult to actually tell which birds were which. Can we use wing and tail length to predict the belly type?

DA Goal 3: Classification – use data on variables from known groups to develop a rule for classifying future observations.

Who uses discriminant analysis ? Everyone!

- Natural Sciences

S.D. Ousley, *Forensic Classification and Biodistance in the 21st Century*, Biological Distance Analysis, 2016

Tomblin, D. C., Adler, G. H. 1998. Differences in habitat use between two morphologically similar tropical forest rodents. *Journal of Mammalogy*, **79**: 953-961

Jordan, M. M., Mateu, J., Boix, A. 1998. A classification of sediment types based on statistical multivariate techniques. *Water Air and Soil Pollution*, **107**: 91-104.

Several students in recent years used DA to distinguish between gender in turtles and ferrets.

- Sociology and Psychology - lots of examples to discriminate between populations, personality types, etc.

Panic disorder subtypes: further clinical differences. Onur, Elif; Alkın, Tunc; Tural, Ümit. *Depression & Anxiety* (1091-4269), 2007, Vol. 24 Issue 7, p479-486, 8p, 2 charts; DOI: 10.1002/da.20253; (AN 26976956)

Mihalovic, Matús. *Performance comparison of multiple discriminant analysis and logit models in bankruptcy prediction*. Economics & Sociology; Ternopil Vol. 9, Iss. 4, (2016): 101-118.

- Medicine –

Samar Hejazi, ... Sheila K. Marshall. *Developmental Trajectories of Weight Status in Childhood and Adolescence*. Global Perspectives on Childhood Obesity (Second Edition), 2019

- IRS – discriminate between audited returns that yield underpayment and those that don't.

D.A. vs. ANOVA *(two sides of the same coin)*

- In D.A., the goal is to find variables that help discriminate between objects of several known group types.
- In (M)ANOVA, the goal is to see if there are differences in the multivariate means between several groups.
- Several tests/techniques used in Discriminant Analysis are also used in MANOVA (I often debate which to teach first) – so some of this material will return in a few weeks.



The Geometry of Discriminant Analysis

Discriminant Analysis produces a weighted linear combination of variables called the **discriminant function**:

$$z = \mathbf{w}'\mathbf{x} = w_1x_1 + w_2x_2 + \dots + w_px_p$$

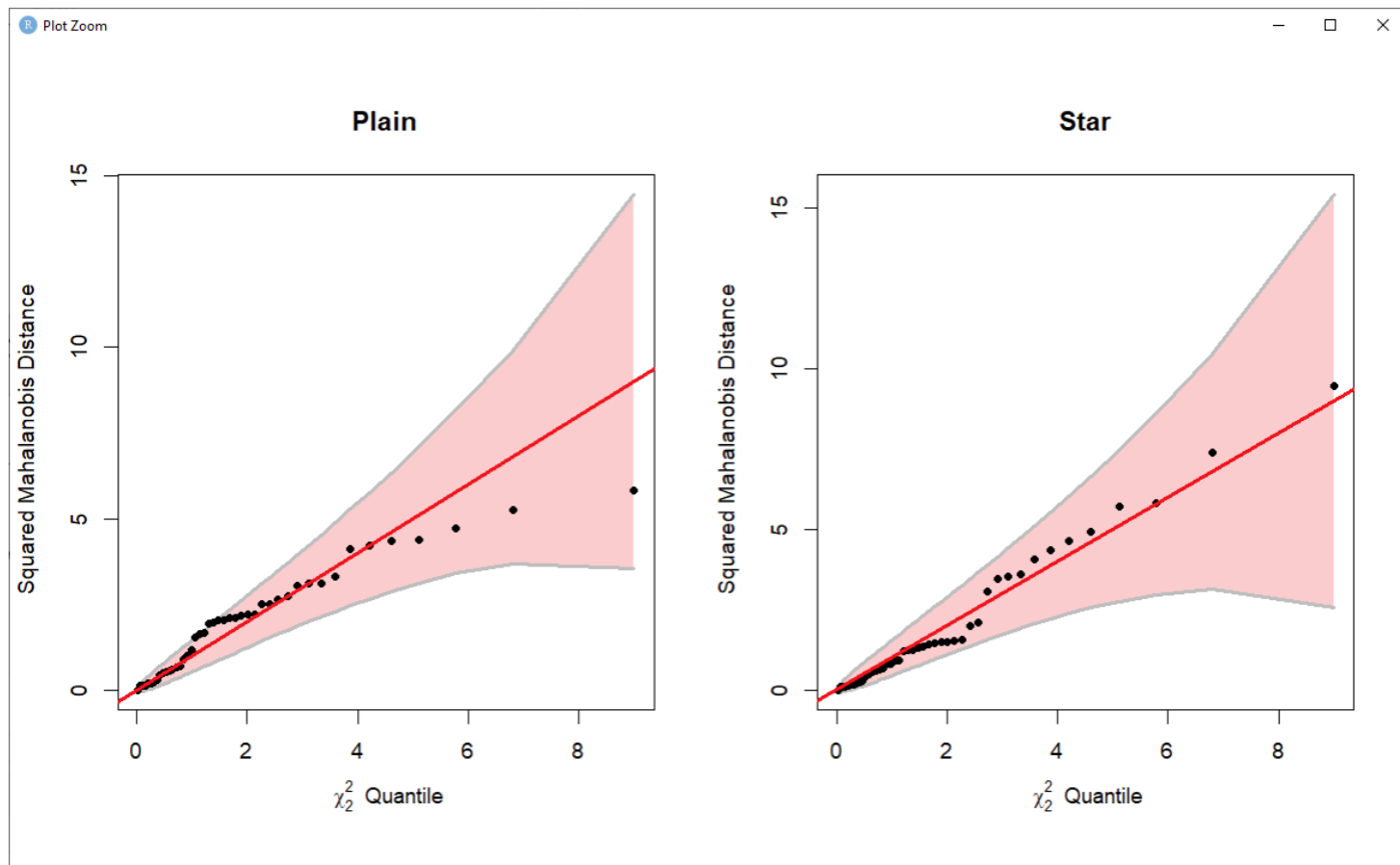
This is the same as what we did in Principle Components!

Discriminant Analysis Assumptions

- 1) The observations **within each group** represent a sample from a **multivariate normal distribution**

This can be checked by making a chi-square quantile plot for the data in each group (not for all data simultaneously)!

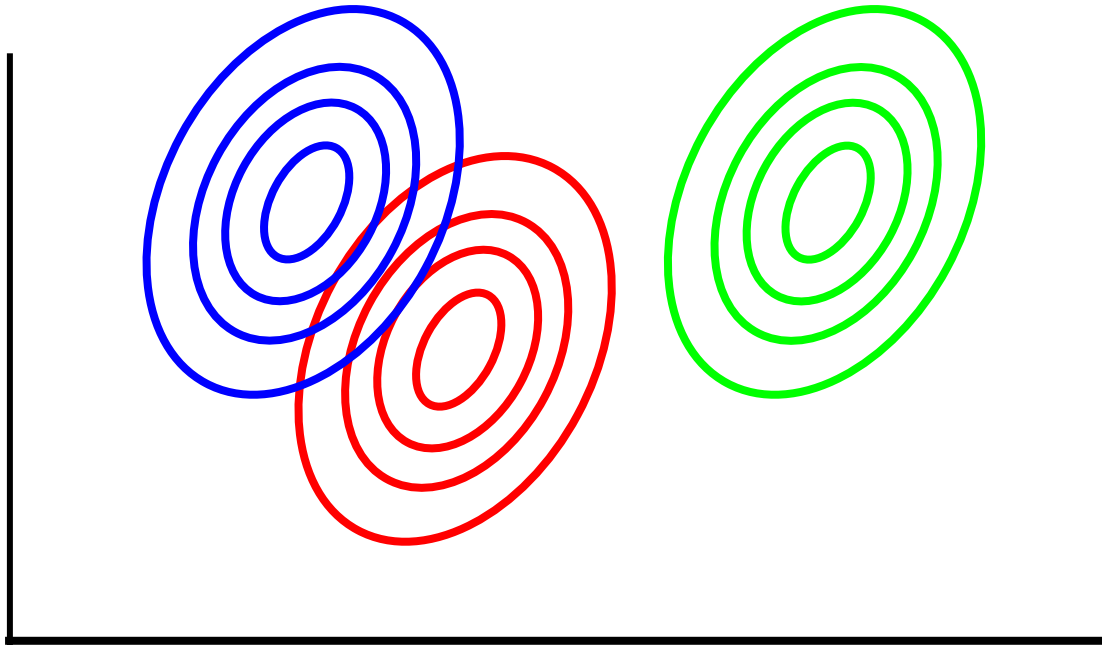
Example: Sneetches. Here is plot for each group:



2) The covariance matrices of each group are assumed to be identical

- Variance of variables is identical between groups
- **Covariance (Correlation)** between variables is **identical** between groups
- This corresponds to having several multivariate normal distributions **which are identical except for the location of their multivariate means.**

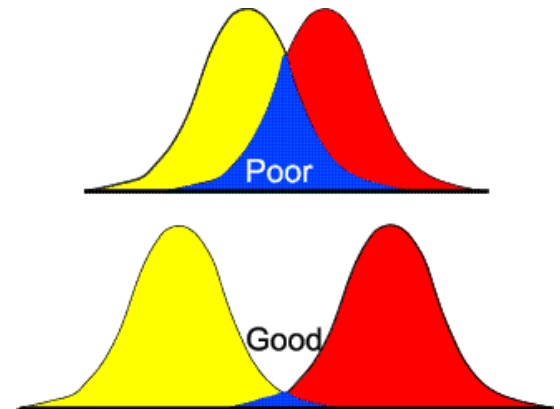
○



More on checking this assumption in a bit . . .

Now:

- Inside of each group, we're assuming data has a multivariate normal distribution. Since z is a linear combinations of normal variables, z has a univariate normal distribution inside each group [*\(link for a proof\)*](#).
- Discriminant scores on the new variable z can be calculated for each group.
- The hope is that the distributions of the scores for each group will be well separated



Goal: Pick weights w_i to maximize the separation of groups on the variable z .

Choosing weights w_i : Sum of Squares of z

- Recall - z is a rotation in p -dimensional space.
- Total variability of \mathbf{X} accounted for by z changes with the rotation (measured by the Total Sum of Squares of z , SS_t).
- In Principle Components Analysis, goal was to find a rotation (i.e. to pick weights) that maximized SS_t .
- In DA, we partition SS_t into Sum of Squares between groups (SS_b) and Sum of Squares within groups (SS_w) (just like ANOVA!):

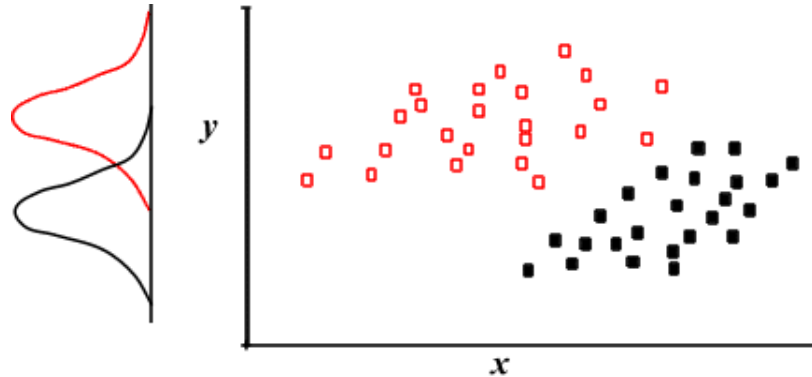
$$SS_t = SS_b + SS_w$$



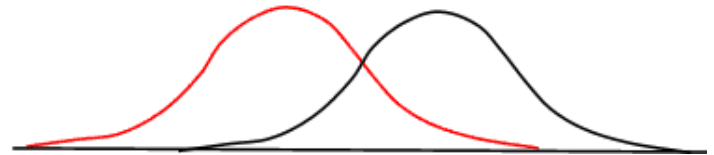
**Goal of DA: choose weights w_i to
maximize the ratio $\lambda = \frac{SS_b}{SS_w}$**

Here's what this looks like for two variables:

- Assume each group has a bivariate normal distribution with identical covariance matrices (i.e. the shape of each group is approximately the same, but the multivariate mean of each group is different).

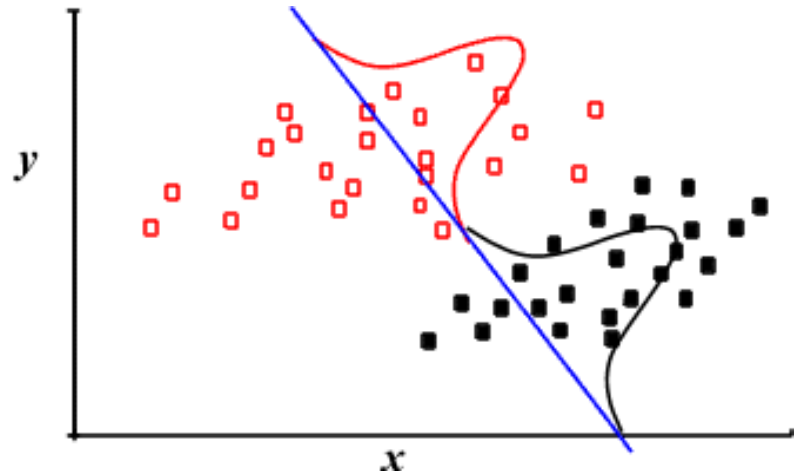


- While there is a separation between groups on each variable, there is still overlap between groups (notice that y gives better separation than x).



- Choose a rotation to maximize the separation between groups – that is

$$\text{maximize } \lambda = \frac{SS_b}{SS_w}$$



*We don't want direction of maximum variability, we want
direction of maximum discrimination!*

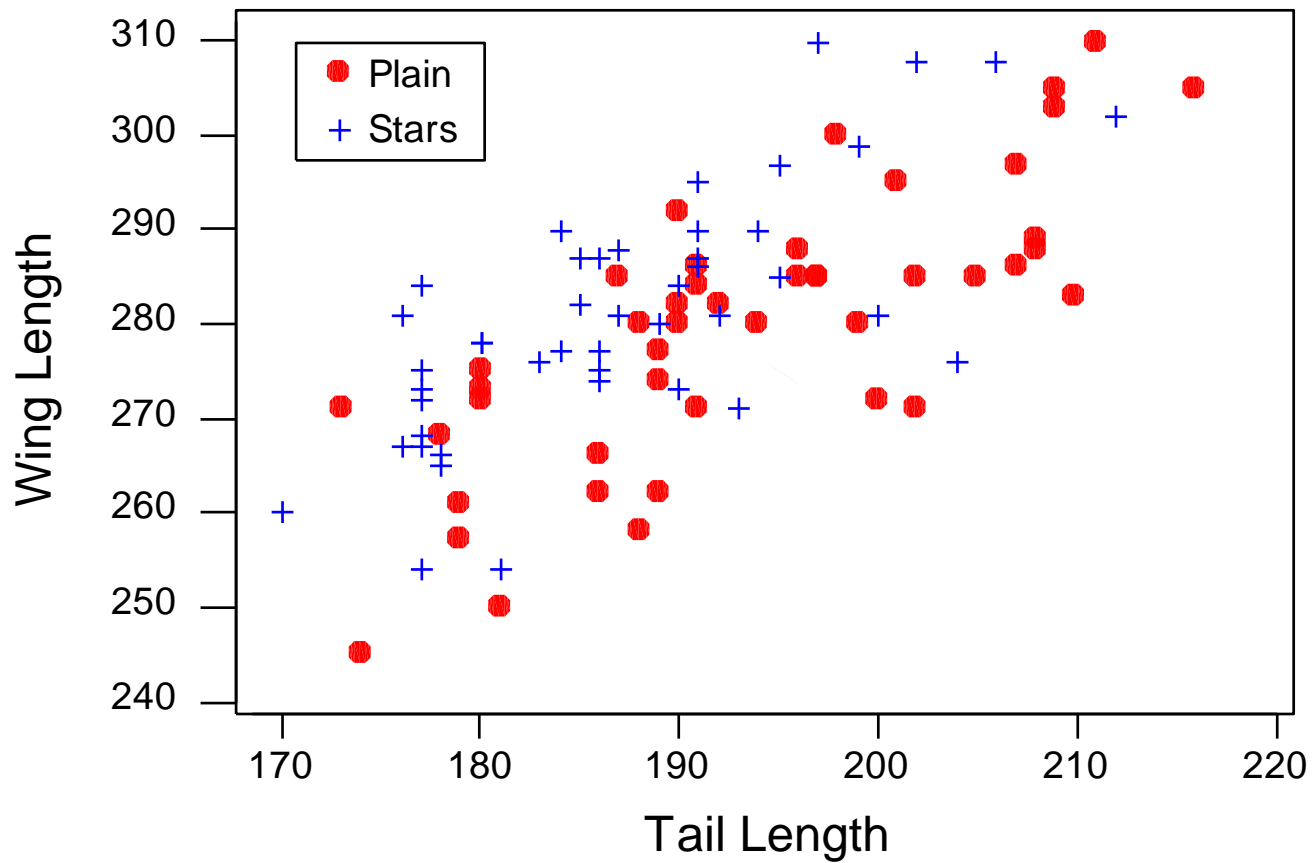
Example: Sneetches. Plot the mean-centered data: the best separation between data appears to be somewhere in direction of arrow, i.e. some rotation between 90° and 180° .



Consider a variety of rotations between 0° and 180° . For each, calculate $\lambda = SS_b / SS_w$, see which rotation makes this the biggest.

Recall that weights \mathbf{W} are calculated based on the angle of rotation θ :

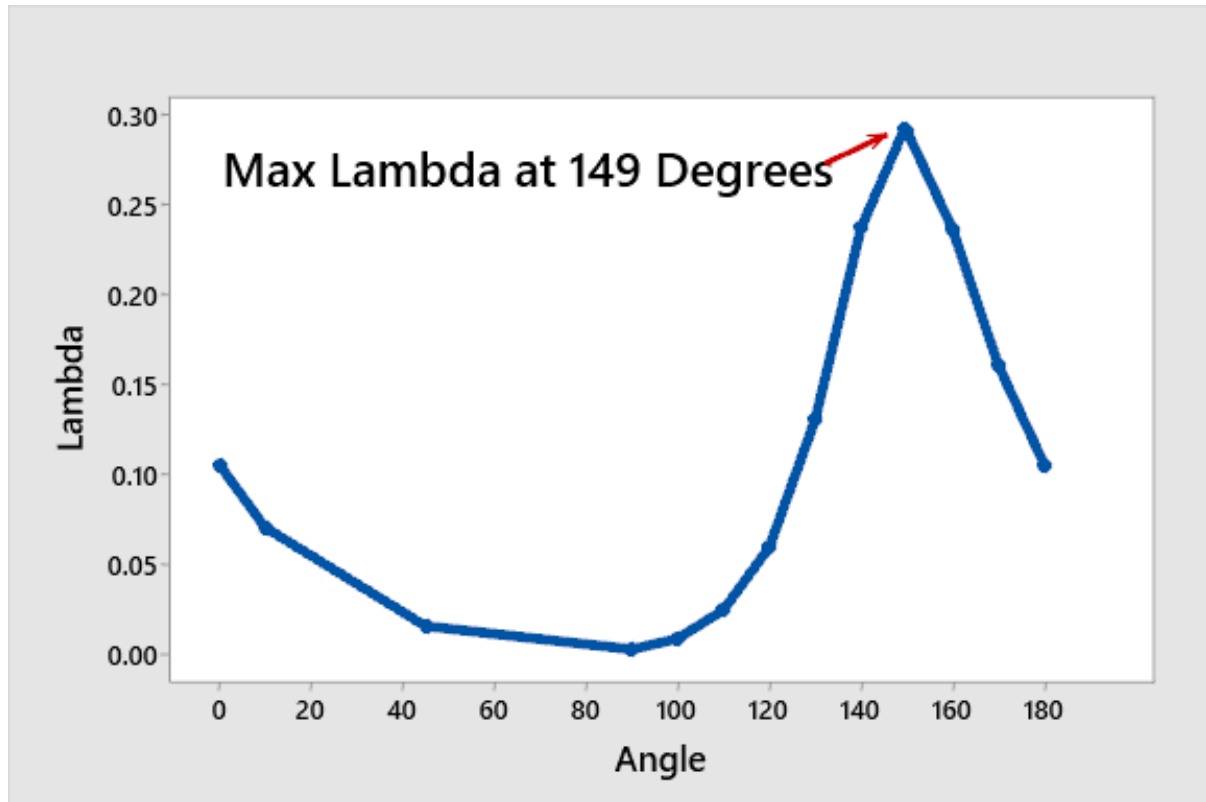
$$z = w_1 \text{Tail} + w_2 \text{Wing} = \cos \theta \text{Tail} + \sin \theta \text{Wing}$$



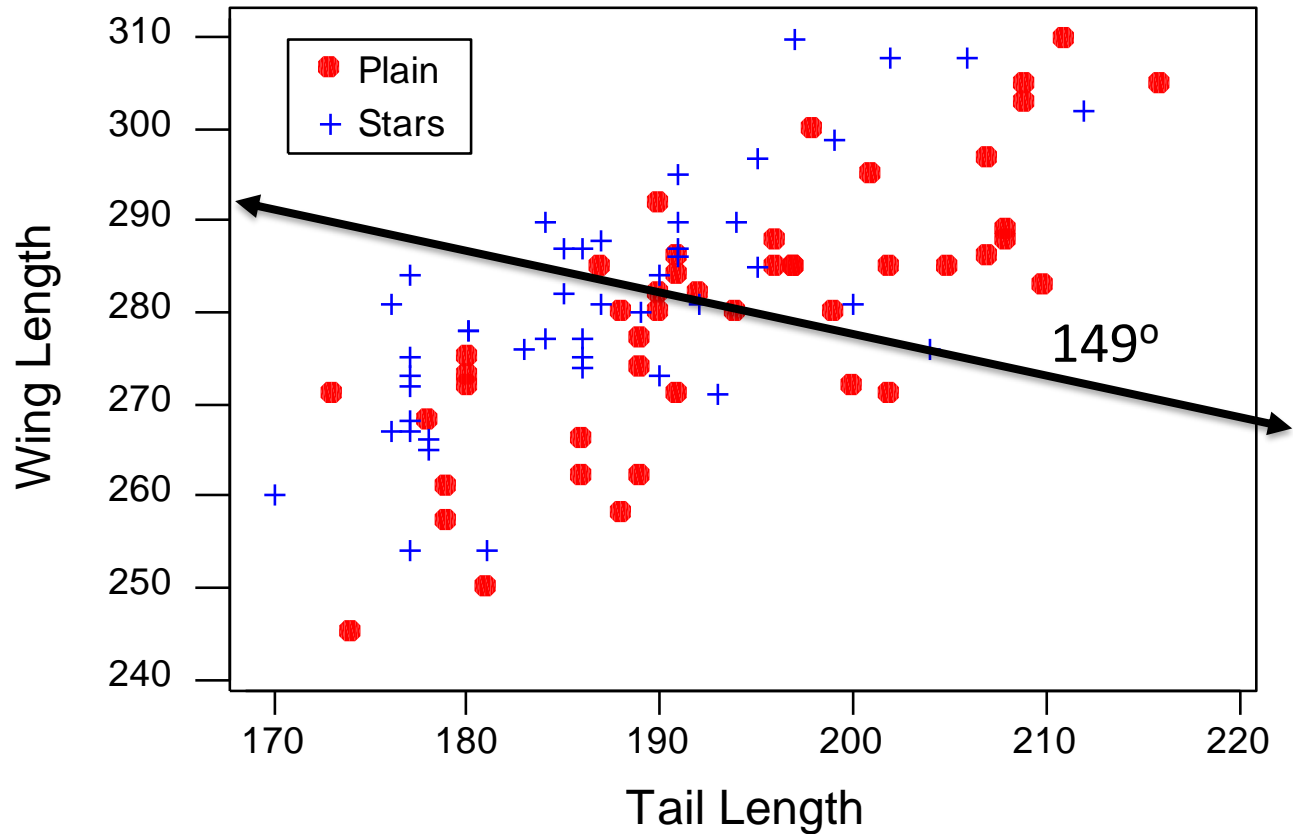
PCA would be a rotation of about 45°

Rotation	<u>weights</u>		<u>Sum of Squares</u>			<u>λ</u>
Angle θ	w_1	w_2	SS_t	SS_w	SS_b	SS_b / SS_w
0	1	0	10081	9120	960	0.105
10	0.98	0.17	13393	12516	877	0.07
45	0.71	0.71	<u>22471</u>	22137	333	0.015
90	0	1	16650	16623	27	0.002
100	-0.17	0.98	13337	13228	110	0.008
110	-0.34	0.94	10028	9789	239	0.024
120	-0.5	0.87	7122	6723	399	0.059
130	-0.64	0.77	4968	4398	570	0.13
140	-0.77	0.64	3827	3095	732	0.237
149.44	-0.86	0.51	3806	2946	859	0.292
150	-0.87	0.5	3837	2972	866	0.291
160	-0.94	0.34	4996	4042	954	0.236
170	-0.98	0.17	7164	6177	987	0.16
180	-1	0	10081	9120	960	0.105

The angle that gives the maximum separation is 149.44° .



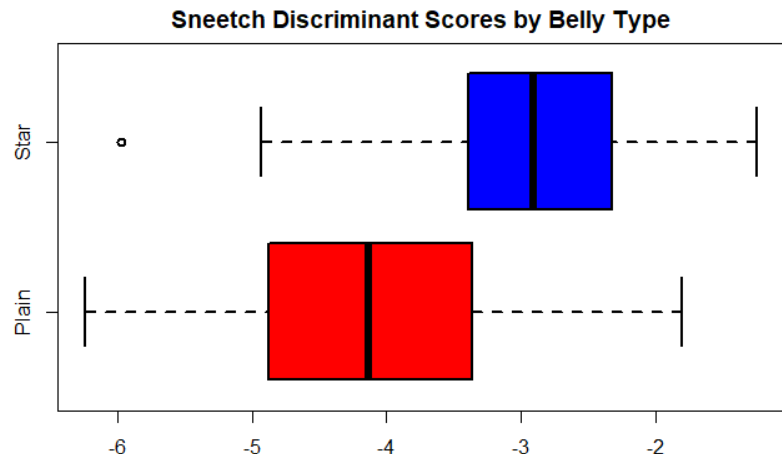
The data with the discriminant function at 149°:



Use the discriminant function to calculate discriminant scores:

$$z = -0.86\textit{Tail} + 0.51\textit{Wing}$$

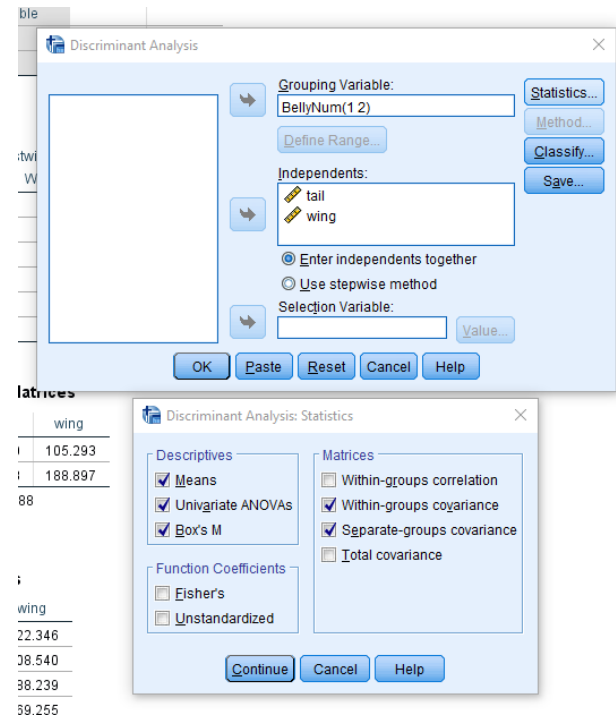
A boxplot of z (i.e. of values projected onto the direction of max discrimination) shows the improved degree of separation between groups (compare to boxplots of original variables)





Computer Notes: In MINITAB, use Stat → Multivariate → Discriminant Analysis. MINITAB is good for DA, but you'll ultimately want to use some other program as well.

For SPSS (which is **AWESOME** for DA) use: Analyze → Classify → Discriminant. From here, enter independent variables. Grouping variable seems to have to be NUMERIC (seems silly!). Ask for group statistics by clicking on Statistics button.



DA in SAS (also *AWESOME* for DA): Use Proc Discrim.



```
PROC DISCRIM DATA=MYDATA.SNEETCH PSSCP TSSCP WSSCP
      BSSCP MANOVA;
  CLASS BELLY;
  VAR TAIL WING;
RUN;
```

Examples of Discriminant Analysis in SAS can be found on the CANVAS home page.

DA in R (*which is OK* for DA): First, load the MASS package. Then use the `lda()` function. There are other packages/functions that I discuss in the posted R code.



Examples of Discriminant Analysis in R can be found on CANVAS.

To continue our discussion of DA, let's look at SPSS output for the Sneetches data: included are some statistics for evaluating our analysis.

Group Statistics					
BellyNum		Mean	Std. Deviation	Valid N (listwise)	
				Unweighted	Weighted
1.00	tail	193.6222	10.98613	45	45.000
	wing	279.7778	14.44093	45	45.000
2.00	tail	187.0889	9.30499	45	45.000
	wing	280.8667	13.00979	45	45.000
Total	tail	190.3556	10.64262	90	90.000
	wing	280.3222	13.67753	90	90.000

These are your basic group summary statistics. Notice that marginal standard deviations in each group are similar (they should be equal if covariance matrices are similar).

Next, SPSS shows us the sample covariance matrices in each group as well as the overall ‘pooled’ covariance matrix:

Covariance Matrices

BellyNum		tail	wing
1.00	tail	120.695	122.346
	wing	122.346	208.540
2.00	tail	86.583	88.239
	wing	88.239	169.255

Pooled Within-Groups Matrices^a

		tail	wing
Covariance	tail	103.639	105.293
	wing	105.293	188.897

a. The covariance matrix has 88 degrees of freedom.

Remember : DA assumes the covariance matrices (the multivariate ‘footprints’) are the same across groups.

Of course, the sample covariance matrices won't be exactly the same. However, here is a good quick rule of thumb:

If the ratio of largest to smallest sample variance/covariance is less than 4, it is reasonable to assume that the covariance matrices are relatively similar.

(think ANOVA rule of ratio of largest/smallest sample SD < 2)

In this case,

Ratios of Largest/Smallest Sample Variance/Covariance

		tail	wing
Covariance	tail	1.4	-
	wing	1.4	1.2

i.e., it's reasonable to assume the covariance matrices are the same in the two belly groups.

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
tail	.905	9.267	1	88	.003
wing	.998	.141	1	88	.708

Interpretation: group means are different for tail length but not for wing length. But – how is this calculated?

Univariate Tests of equality of group means

Usual thing – do a *t*-test to see if means are different.



So, what is Wilks' Lambda, and why is there an F-statistic?

Recall the symmetric, dimension- p matrix of **total sums of squares and cross-products for MEAN CENTERED DATA**:

$$\text{SSCP}_t = \mathbf{X}'_d \mathbf{X}_d = \begin{bmatrix} s_{11} & s_{21} & \Lambda & s_{p1} \\ s_{12} & s_{22} & \Lambda & s_{p2} \\ \text{M} & \text{M} & \text{O} & \text{M} \\ s_{1p} & s_{2p} & \Lambda & s_{pp} \end{bmatrix}$$

These are the matrices of mean-centered data (subtract mean from each column – here d stands for ‘deviations’)

$$\text{where } s_{jk} = \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$

Notationally \bar{x}_j, \bar{x}_k are the means of the j 'th and k 'th variables, i.e. $j, k \in 1, 2, \dots, p$

That is: the diagonal elements are the total sum of squares for each individual variable, and the off-diagonals are sums of cross products between variables.

For the Sneetch data:

$$\mathbf{SSCP}_{\substack{2 \times 2 \\ \text{Total}}} = \mathbf{X}'_{\substack{2 \times 90d}} \mathbf{X}_{\substack{90 \times 2d}} = \begin{bmatrix} SS_{Tail} & CP_{TailWing} \\ CP_{TailWing} & SS_{Wing} \end{bmatrix} = \begin{bmatrix} 10080 & 9105 \\ 9105 & 16650 \end{bmatrix}$$

*Note that if you divide by $(n - 2)$
 $= 88$, you get the **Sample
 Covariance Matrix:***

$$\mathbf{S} = \frac{\mathbf{SSCP}_{Total}}{88} = \begin{bmatrix} 114 & 103 \\ 103 & 189 \end{bmatrix}$$

*AND: if you take the square-root of \mathbf{S} ,
 the diagonal entries give the **Standard
 Deviations** for each variable*

$$\sqrt{\mathbf{S}} = \begin{bmatrix} 10.6 & \\ & 13.7 \end{bmatrix}$$

NOW: Define an equivalent matrix for **within group sums of squares and cross-products**, SSCP_w and **between group sums of squares and cross-products** SSCP_b (based on the group means).

$$\text{No surprise: } \text{SSCP}_t = \text{SSCP}_w + \text{SSCP}_b$$

Let's see how SAS does discriminant analysis:

Discriminant Analysis in SAS: Use Proc DISCRIM. I'm including options here to get the **SSCP** matrices:



```
PROC DISCRIM DATA=MYDATA.SNEETCH PSSCP TSSCP WSSCP  
              BSSCP MANOVA;  
  CLASS BELLY;  
  VAR TAIL WING;  
RUN;
```

The results are below. Notation:

$G = 2$ groups

\bar{x}_{1k} = mean of the k th variable in the 1st group (i.e. mean tail length in the plain belly group)

$j, k \in 1, 2, \dots, p$ are indices of the variables (i.e. tail, wing)

Sum of squares within the Plain Belly Group – how far are observations from the group mean for each variable (and cross products)?

Variable	Belly = Plain		
	Label	tail	wing
tail	tail	5310.57	5383.22
wing	wing	5383.22	9175.77

SSCP_{Within(Group=1)}

$$s_{jk} = \sum_{i=1}^{n_1} (x_{ij} - \bar{x}_{1j})(x_{ik} - \bar{x}_{1k})$$

Sum of squares within the Star Belly Group – how far are observations from the group mean for each variable (and cross products)?

Variable	Belly = Star		wing	SSCP _{Within(Group=2)}
	Label	tail		
tail	tail	3809.64	3882.53	$s_{jk} = \sum_{i=1}^{n_2} (x_{ij} - \bar{x}_{2j})(x_{ij} - \bar{x}_{2k})$
wing	wing	3882.53	7447.20	

Sum of squares within the Star Belly Group **AND** the Plain Belly Group – how far are observations from the group mean for each variable **SUMMED OVER GROUPS?**

Pooled Within-Class SSCP Matrix				SSCP _{Within}
Variable	Label	tail	wing	
tail	tail	9120.22	9265.75	$s_{jk} = \sum_{g=1}^G \left(\sum_{i=1}^{n_g} (x_{ij} - \bar{x}_{gj})(x_{ij} - \bar{x}_{gk}) \right)$
wing	wing	9265.75	16622.97	

Sum of squared distances **between group means and overall mean** for each variable (and cross products)?

Between-Class SSCP Matrix			
Variable	Label	tail	wing
tail	tail	960.40	-160.06
wing	wing	-160.06	26.67

SSCP_{Between}

$$s_{jk} = \sum_{g=1}^G n_g (\bar{x}_{gj} - \bar{x}_j)(\bar{x}_{gk} - \bar{x}_k)$$

Total sum of squared distances from observations to overall mean (and cross products)

Total-Sample SSCP Matrix			
Variable	Label	tail	wing
tail	tail	10080.62	9105.68
wing	wing	9105.68	16649.65

SSCP_{Total}

$$s_{jk} = \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$$

Wilks' Lambda (Λ) (*Wilks, 1932*): for each variable,

$$\Lambda = \frac{SS_w}{SS_t}$$

(just read diagonal elements of \mathbf{SSCP}_t and \mathbf{SSCP}_w)

Factoid: Wilks' Lambda can be converted to an F-distribution
(see *Johnson and Wichern, p. 300* for more details or read [HERE](#)):

$$F = \left(\frac{1 - \Lambda}{\Lambda} \right) \left(\frac{n - p - 1}{p} \right)$$

Another Factoid: the square of a t -statistic has an F distribution

SO : Wilks' Lambda is equivalent to doing the usual t-test for difference of means in two groups!

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
tail	.905	9.267	1	88	.003
wing	.998	.141	1	88	.708

i.e. Wilks' Lambda shows that there are differences between bellies in tail length, but not in wing length

Then why didn't you just use the !!@! t-test!!?!!?!!*



Summary of Canonical Discriminant Functions

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.774	22.263	2	.000

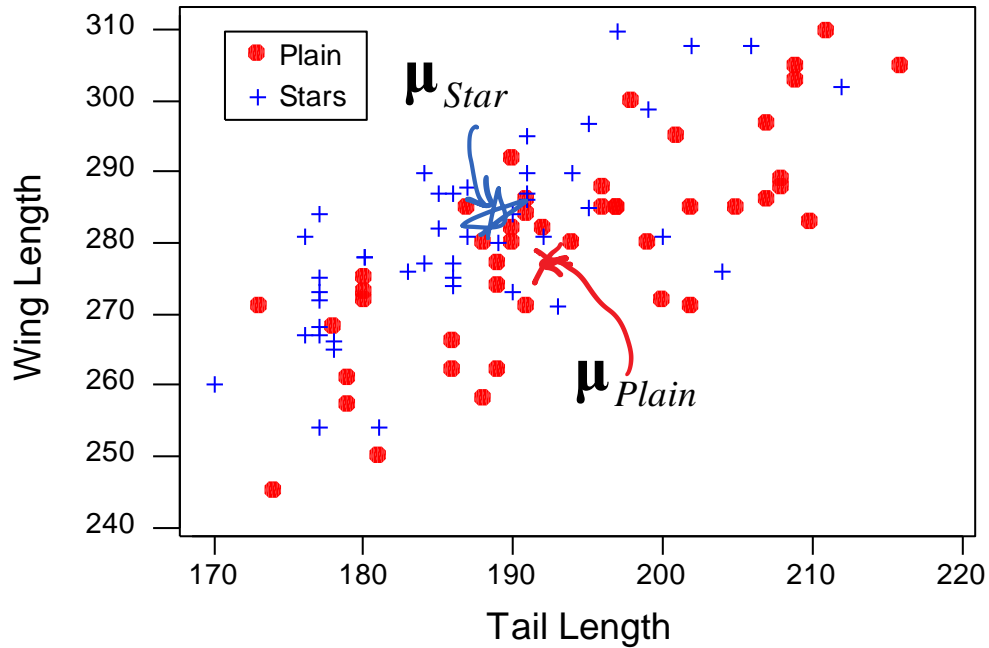
Rather than simply testing for different means between groups for each individual variable, we'd like to **test for differences between groups for all variables simultaneously** – i.e. we want a

Multivariate Test of Group Means

$$H_o : \mu_{Plain} = \mu_{Star} \quad H_a : \mu_{Plain} \neq \mu_{Star}$$

i.e.

$$H_o : \begin{pmatrix} \mu_{Plain}^{Tail} \\ \mu_{Plain}^{Wing} \end{pmatrix} = \begin{pmatrix} \mu_{Star}^{Tail} \\ \mu_{Star}^{Wing} \end{pmatrix} \quad H_a : \begin{pmatrix} \mu_{Plain}^{Tail} \\ \mu_{Plain}^{Wing} \end{pmatrix} \neq \begin{pmatrix} \mu_{Star}^{Tail} \\ \mu_{Star}^{Wing} \end{pmatrix}$$



Define the Multivariate version of **Wilks' Lambda**:

$$\Lambda = \frac{|\mathbf{SSCP}_w|}{|\mathbf{SSCP}_t|} \text{ where } | \quad | \text{ is the determinant}$$

Under the **null hypothesis of no difference in multivariate means**, Wilks' Lambda should have an approximately chi-square distribution ([link](#)):

$$\chi^2_{p(G-1)} = - \left[n - 1 - \left(\frac{p + G}{2} \right) \right] \ln \Lambda$$

where G is the number of groups.

SO: we reject the hypothesis that the multivariate means are equal between bellies!



If this seems a bit cursory, don't worry – it's all coming back in MANOVA!

Eigenvalues

Function	Eigenvalue	Eigenvalues		Canonical Correlation
		% of Variance	Cumulative %	
1	.292 ^a	100.0	100.0	.475

a. First 1 canonical discriminant functions were used in the analysis.

The computer solution to find the rotation that gives best discrimination is another eigenvalue problem: pick weights to maximize

$$\lambda = \frac{\mathbf{w}' \mathbf{SSCP}_b \mathbf{w}}{\mathbf{w}' \mathbf{SSCP}_w \mathbf{w}}$$

Taking partial derivatives with respect to \mathbf{w} and setting to zero, we have a maximum when

$$(\mathbf{SSCP}_w^{-1} \mathbf{SSCP}_b - \lambda \mathbf{I}) \mathbf{w} = 0$$

i.e., an eigenvalue problem with associated eigenvector \mathbf{W} .

- For two groups, there is only one eigenvalue, here equal to .292. The corresponding eigenvector points in **the linear direction of maximum separation**.

Equality of Covariance Matrices: An Assumption of DA!!

Box's Test of Equality of Covariance Matrices

Log Determinants		
BellyNum	Rank	Log Determinant
1.00	2	9.230
2.00	2	8.835
Pooled within-groups	2	9.047

The ranks and natural logarithms of determinants printed are those of the group covariance matrices.

Test Results		
Box's M		1.253
F	Approx.	.407
	df1	3
	df2	1393920.000
	Sig.	.748

Tests null hypothesis of equal population covariance matrices.

Remember that DA assumes

- Data from both groups comes from a multivariate normal distribution (test with Q-Q Chi-square)
- **Covariance matrices in each group are identical (i.e. same shapes, different multivariate means / centroids)**

We already examined univariate variances for individual variables in each group

Box's M test for equality of entire Covariance matrices
([Link](#)):

$$H_o : \Sigma_1 = \Sigma_2 = \dots = \Sigma_G$$

$$H_a : \text{Not all of } \Sigma_i \text{ are equal}$$

- Basically compares how far the individual sample Covariances matrices S_i for each group are from a pooled estimate of the overall Covariance matrix S_{pool} (think about pooled estimates of the standard deviation s in ANOVA).
- Based on log of the determinants of these matrices:

$$M = (n-G)\ln|\mathbf{S}_{pool}| - \sum_{j=1}^G (n_j - 1)\ln|\mathbf{S}_j|$$

- Box's M can be approximated by an F-test (better – used by SPSS) or a Chi-square test (ok, used by R) – if significance is low, reject hypothesis of equality of covariance matrices (which is unfortunate for us!)

HOWEVER – Box's M tends to be overly sensitive to even moderately large sample sizes and minor deviation from equality (rejects hypothesis too easily) **and** is also sensitive to departures from non-normality (i.e. we might be testing shape rather than multivariate spread).



SO –

- 1) Look at raw log(determinants) – see if they are approximately equal.
- 2) Use a lower than usual threshold – i.e., instead of .05, use .001 as a threshold for rejecting null hypothesis in Box M.

Note – *Box's M Test is available in SPSS, R (see online Code). Not easily available in SAS (would take some coding) .*

..



Example: Sneech Data – no evidence that covariance matrices are not similar – p -value is large and raw log determinants are relatively similar. That is, we fail to reject the hypothesis of unequal covariance matrices, so we proceed assuming they are the same (not a legitimate hypothesis test, but does provide guidance!)

Box's Test of Equality of Covariance Matrices

Log Determinants

BellyNum	Rank	Log Determinant
1.00	2	9.230
2.00	2	8.835
Pooled within-groups	2	9.047

The ranks and natural logarithms of determinants printed are those of the group covariance matrices.

Test Results

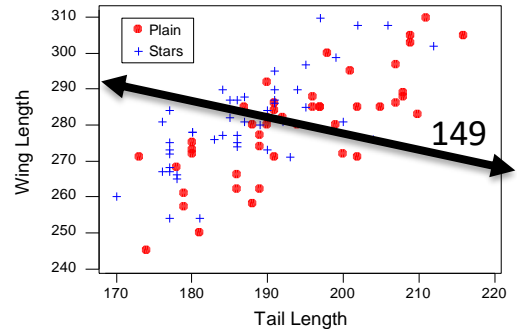
Box's M		1.253
F	Approx.	.407
	df1	3
	df2	1393920.000
	Sig.	.748

Tests null hypothesis of equal population covariance matrices.

Discriminant Function Coefficients

Coefficients of the Discriminant Function are unique only up to a constant:

- They specify the **direction** of the maximum discrimination, **not the magnitude**.
- However, the **ratios of coefficients** are unique for unstandardized data.



Unstandardized Coefficients *(FYI, this is what is reported in R)*

- Coefficients calculated based on the unstandardized data
- Also called the **canonical discriminant function** (special case of canonical correlation – coming later)
- **Constant simply makes average of discriminant scores equal to zero.**

Canonical Discriminant Function Coefficients

	Function 1
tail	.149
wing	-.088
(Constant)	-3.560

Unstandardized coefficients

Hey – these are not the coefficients we calculated ‘by hand’ using geometry . . .

Normalized Coefficients

- Normalize unstandardized coefficients so **sum of squared weights equals one** (*i.e. make the vector one unit long!*)
- Divide each unstandardized coefficient by

$$\sqrt{w_1^2 + w_2^2 + \dots + w_p^2}$$

For Sneeck Data, $\sqrt{(.149)^2 + (.088)^2} = .017$

So coefficients are

$$.149 / .017 = 0.86, \text{ and } -.088 / .017 = -0.51$$

which (to a factor of minus one) are what we found before.

THIS IS THE ASSOCIATED EIGENVECTOR!

Standardized Coefficients

- Coefficients based on **standardized data** (i.e. make variables have s.d. 1)

- Often used to assess the **relative importance of discriminator**

variables: variables with larger standardized coefficients are considered to be 'better' discriminators

- **HOWEVER:** if variables are correlated with each other, use caution – same problem as in multiple regression – variables are competing to discriminate between groups.

Standardized Canonical Discriminant Function Coefficients

Function 1	
tail	1.514
wing	-1.214

Appears that Tail is a better discriminator than Wing (suggested by boxplots) – however, correlation between variables is 0.7 – definitely correlated so be careful about making inferences about relative discriminating power.